Determination of iron by Z-GFAAS and the influence of short-term precision and long-term precision

Andersen, Jens

Published in:
Central European Journal of Chemistry

Link to article, DOI:
10.2478/s11532-013-0362-9

Publication date:
2014

Cent. Eur. J. Chem.
DOI -

Central European Journal of **Chemistry**

Determination of iron by Z-GFAAS and the influence of short-term precision and long-term precision

Research Article

Jens E.T. Andersen*

Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Received 2 May 2013; Accepted 27 September 2013

Abstract: A detailed method validation of graphite-furnace atomic absorption spectrometry (GFAAS) with Zeeman background correction was performed. The aim is to perform a detailed investigation of short-term precision as opposed to long-term precision. It was suggested that release of graphite flakes into the light path during measurement significantly influenced the performance of the method. It was found that significant deviations with respect to the certified values were frequent and an estimate of reliable uncertainties was obtained only after a high number of repetitions. Uncertainty of *intralaboratory* testing was evaluated as a method to estimate uncertainties that are comparable to uncertainties that were obtained by ~~*intralaboratory*~~ testing and to uncertainties predicted by the Horwitz curve. To a large extent, the uncertainty in measurement that was predicted by pooled calibrations corresponded to the uncertainties that were obtained from multiple determinations of unknowns. It was thus proposed that a large proportion of the difference in uncertainty in measurement between laboratories could be explained by properties of the different detectors. In order to support accuracy, it is suggested that a higher level of uncertainty should be accepted in analytical investigations.

Keywords: *Quality assurance • Iron analysis • Z-GFAAS • Z-ETAAS • Method validation*

1. Introduction

Results of interlaboratory comparisons show that professional laboratories cannot agree about the average value nor the uncertainty in measurement of a certified reference material (CRM) [1]. Since results of analytical chemistry are of paramount importance to decision making [2] in life sciences and industry, it follows that a consensus must be attained about practices and procedures in laboratories. In a survey of the International Measurement Evaluation Programme (IMEP), Bednarova *et al*. [3] reported large deviations documented in ~~te~~ determination of trace elements in water. Earlier interlaboratory comparisons by the IMEP showed deviations by more than an order of magnitude between reported average values with associated uncertainties which also deviated by more than an order of magnitude [4–6]. Although some measurements were performed by laboratories using different types of analytical technologies [4,6], there was a general trend of estimating significantly different values when the measurements were performed by laboratories using the same technology, *e.g.* for analysis of cadmium and lead in water, as shown by Papadakis *et al*. [5], standard deviations of the mean, confidence ranges [7], combined standard uncertainties or expanded uncertainties [8]. In principle it is acceptable to report the uncertainty of measurement in terms of standard deviations of the mean or confidence ranges however the number of repetitions must be reported. In order to facilitate comparison and evaluate correctly the uncertainty in measurement, Eurachem [2,9] decided in 2008 that uncertainty in measurement should always be reported as expanded uncertainties that are calculated upon the basis of the Guide to the expression of Uncertainty in Measurement (GUM) [8] and the guide to Quantify Uncertainty in

* *E-mail: jeta@kemi.dtu.dk*

Table 1. Figure of merits that are associated with determination of iron by electrothermal AAS (ETAAS)/GFAAS (Ordinary wavelength 248.3 nm).

Technology (Wavelength in nm)	Matrix	LOD ($mg \cdot dm^{-3}$)	Linear range ($mg \cdot dm^{-3}$)	Sensitivity ($10^{-3} abs \cdot mg^{-1} \cdot dm^3$)	Intercept (10^{-3}Abs)	Injection volume (mm^{-3})	RSD (%)	Ref.
ET (248.3)	Human breast milk	1.4	0 - 30	4.48	5.0	20	< 10	[16]
ET (248.3)	Certified reference	0.6	0 – 25	2.63	10.3	20	< 4	This work (Precision)
GF (242.8)	Rain water	5	0 - 6	-	-	20	2.4	[24]
GF (248.3)	Bovine milk	0.39 – 0.60	-	-	-	10	< 7	[7]
GF (248.3)	Multi-element sample	0.2*	-	-	-	-	2.4 – 3.5	[25]
GF (248.3)	Caco-2 cells	1.3	-	-	-	-	-	[17]
GF (248.3)	Atmo-spheric water samples	0.2	0 -10	52	30	20	7.8	[26]
GF (248.3)	Human hair	$3.2 \cdot 10^{-4}$	-	-	-	10	< 8	[18]
GF (248.3)	LaF_3	3	0 - 150	-	-	10	-	[27]
GF (248.3)	Rice	0.9	0 - 500	1.5	54.6	-	-	[28]
GF (372)	Human serum	5.6**	0 – 14	21.6	-16.2	25	-	[19]
ET (248.3)	Cork stoppers	0.8	0.8 - 50	-	-	25	1.4	[29]
GF	Bleomycin	1.12	-	-	-	-	2.2	[30]
ET (248.3)	Legumes and nuts	0.2*	-	0.3	-	-	4.1 – 4.6 (Precision)	[31]
GF (248.3)	Butter	-	0 – 25	7	0/6	10	7 – 10%	[32]

**Estimated by conversion from w/w to $mg \cdot dm^{-3}$ using an injection volume of 20 mm^{-3}. **With preconcentration*

Analytical Measurements (QUAM) published by the European Analytical Chemistry (Eurachem) and Cooperation on International Traceability in Analytical Chemistry (CITAC) [10] organisations. Eurachem intends to implement the principles of GUM/Eurachem/CITAC guides to the work of laboratories in the industry where the guidelines of ISO 5725 [11] are still the more important tools of statistical data processing. A correct estimate of uncertainty in measurement is the overall aim for evaluating practices and procedures of analytical chemistry. The issue was illustrated in a comprehensible manner by de Bievre [12] who showed that two laboratories tend to disagree on the value found in each laboratory when the uncertainty is low but an agreement might be obtained if uncertainties were high. Accordingly, a reliable value of the uncertainty is crucial for deciding if results differ or correspond to each other.

The present report is motivated by an interest in understanding the origin of differences displayed in the results and overview of interlaboratory testing [13]. In the publication by de la Calle *et al*. [13] it was clearly shown that professional laboratories were able to agree neither on average values nor on the level of uncertainty. Deviations as large as an order of magnitude of the average value and more than an order of magnitude of the uncertainty demonstrate the extent of the challenges to understanding analytical problems [1,13]. In all the interlaboratory testing, the collected results of individual laboratories constitute a characteristic distribution of results that is reminiscent of an S – shaped curve [1,3–6,13,14]. This indicates that each measurement constitutes part of a distribution related to a consensus value and a corresponding total overall level of uncertainty. Two tentative explanations may be inferred from the problem; either the performance differs between laboratories or the performance is equal but the applied methodology does not allow to discriminate. It may be anticipated that the former suggestion prevails owing to differences in analytical skills and between applied apparatus. Thus, a thorough study was undertaken to determine the uncertainty in measurement related to the determination of iron by graphite-furnace atomic absorption spectrometry (GFAAS) with Zeeman-background correction. Determination of elements by Zeeman GFAAS is characterized by high precision, low limit of detection (LODs) and low consumption of samples. Typically, volumes of 10 – 20 microliters are suitable for analysis of elements in biological samples (Table 1) [15–19]. According to the manufacturer, LODs far below 1 $mg\ dm^{-3}$ are commonly found for metallic

elements but it has also been recognized that matrix effects are important, which promote application of matrix modifiers for better accuracy. These general remarks are reflected in the literature results of Table 1 where large variations were found in figures of merits among different types of samples. On average, the LOD was 1.5 mg dm^{-3} equivalent to a limit of quantitation (LOQ) of 5.0 mg dm^{-3} with an expected relative standard deviation (RSD) of samples much below 10% (Table 1). On the short-term time scale the contribution from the detector is included through the repetition term but long-term variations of the detector have also been proposed to give an even larger contribution to the uncertainty. It was recently shown that these long-term fluctuations are crucial for describing accuracy of the methods ultra-violet-visual (UV-VIS) spectrophotometry [20], flame atomic absorption spectrometry (FAAS) [21,22] and GFAAS [22]. The present work investigates the performance of GFAAS with Zeemann-background correction that may exhibit a dependence of RSD as a function of concentration different from that of conventional GFAAS [23].

2. Experimental procedure

2.1. Standards and samples

The standard solutions were prepared by dilution of solutions certified to 999±2 mg dm^{-3} of iron nitrate in 0.5 M nitric acid (Merck, CertiPUR, $Fe(NO_3)_3$). All solutions were diluted with distilled water. Blanks were prepared as reagent blanks of 0.1 M nitric acid. Four certified reference materials (CRMs) were applied to analysis of iron: (1) NIST SRM 1640 with consensus value of 34.3 ± 1.6 mg kg^{-1}, (2) Sea water lot# 605806 with consensus value of 20 mg dm^{-3}, CRM Waste Water EnviroMAT EU-H-1 (SCP Science): 0.65 mg kg^{-3} at dilution factor (DF) 50; confidence of 0.62-0.68 mg kg^{-3}, CRM Drinking Water EnviroMAT EP-H-1 (SCP Science): 1.48 mg kg^{-1} at DF 100; confidence interval of 1.43-1.53 mg kg^{-3}. $Mg(NO_3)_2$ (Merck) was used as modifier. No information was available about the number of repetitions, and an unknown number of outliers were removed from the data set according to the certificates. All solutions of CRMs were diluted to concentrations that resided within the linear range of responses.

2.2. Apparatus

The measurements were performed by GFAAS (Perkin-Elmer Analyst 600, Zeeman background correction) with an iron-hollow cathode lamp (S&J Juniper) radiation source using the wavelength at 248.3 nm, a cathode current of 40 mA and a slit width of 0.2 nm. The limits of statistics defined in the software were adjusted to allow all measurements to be accepted. A non-zero intercept was frequently found in the regression lines, but the apparatus was not adjusted to zero baselines before determination of unknowns. Aliquots of 20 mL were injected into the graphite tube at room temperature, followed by three steps of drying; first step: ramped for 5 s to 90°C and held for 10 s, second step: ramped for 10 s to 120°C and held for 10 s and third step: ramped for 5 s to 180°C and held for 20 s. Then followed the pyrolysis step where the temperature was ramped from 180°C to 1400°C in 10 s and held there for 20 seconds. After the pyrolysis step, the data were recorded in 5 s during the atomization step at 2400°C. Finally the graphite tube was cleansed by ramping temperature from 2400°C to 2500°C in 5 s and held at 2500°C for 3 s. The internal-flow rate of argon was 250 mL min^{-1}.

3. Results and discussion

3.1. Treatment of outliers

In the present investigation, an outlier is defined as a measurement that may be identified by means of statistics and it is found to deviate significantly from the average value of the distribution at a given level of confidence. In the present investigation several outliers were identified by guard bands [2], Dixon's Q-testing [24] or by Cochran's testing [11] during the course of data processing but none of the outliers thus determined were removed from the data set. All data of calibrations were retained and were included in the interpretation of data. A high number of repetitions were applied to the investigations, and outliers occurred only rarely, as expected from common considerations of probabilities that were derived from distributions. An outlier has only a limited influence on the average value determined by multiple repetitions, and it also has a limited influence on the standard deviation (SD). However, in the case of a low number of repetitions, the outliers may have a tremendous influence on both average values and the level of uncertainty. The identification of an outlier resulted in more repetitions, which diminished the effect of its presence on the final result.

3.2. Errors

Only gross errors were considered. Such errors included incorrect sample preparation, incorrect handling of apparatus and accidents in the laboratory. Errors have got little to do with statistics, and errors cannot be identified unequivocally by means of statistics, as a

rule. Statistics may be applied to identify an outlier that may result from an error but unless a gross error was detected the outlier should be retained and included in calibrations and calculation of concentrations. A result that deviates significantly from the mean value must be retained in the data set and included in the estimate of the final result. It is not supposed that errors appear in the final data set but they may be removed from the data set when the origin of the error is documented and reported in the laboratory journal.

3.3. Single-line calibrations for precision

3.3.1. LOD

The LOD is one of the important figures of merits that informs about the prime performance of the apparatus with respect to one particular type of measurement. For the determination of the LOD, it is common practice to select the highest sensitivity and the lowest noise level from several sets of experiments because the result thus obtained signifies the ultimate performance of the methodology. A LOD value obtained in this manner provides a low concentration that is rarely accessible to reliable analysis because the uncertainty in the measurement is high in the vicinity of the LOD. Several methodologies are available for determination of the LOD and they all provide low concentrations that relate to the inherent noise level of the apparatus. Huber states that the LOD and its uncertainty should be as small as possible, as deduced from three methods for determination of the LOD [25]. In an initial series of experiments, the LOD(3s) [26] was determined by repetition of ten blanks, which yielded large positive values for the areas (Table 2). A surprisingly high value of the area of approx. 2.8 A s was found in one of the measurements but no errors were retrieved during the experiment. The mean value of the results was 0.34 and the SD was 0.88, and it was used together with the slope of Fig. 1 to determine the LOD of 970 mg kg^{-1} that is by far out of scale (Table 2), as compared to results obtained for the calibration (Table 1). The LOD thus determined may not be reliable but, since no errors were detected, ought to be retained as a genuine result of the measurement. However, the experiments that were used to determine this very high LOD value motivated further investigations. Tentatively, a more usual value of the LOD could be determined by performing more repetitions, which might produce fewer outliers and thus provide a more realistic LOD, according to Table 1 with literature values [27]. As mentioned above, it is recommended not to remove outliers from the data set, because it might reduce the chance of retrieving important information about the applied methodology. The LOD determined above (Table 1) has a value that

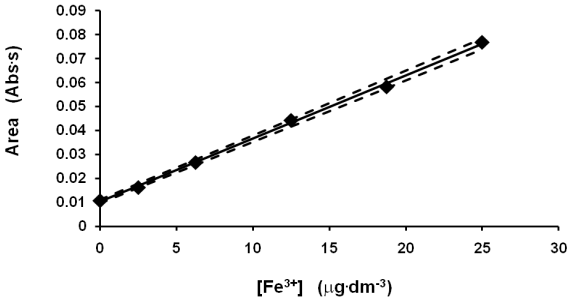

Figure 1. Data of determination of iron by Zeeman GFAAS shown together with the appropriate regression line (full line) and lines representing confidence ranges (95% level of confidence, $p < 0.05$). The slope of the regression line is 0.00263(3) abs s μg^{-1} dm^3 and the intercept is 0.0103(4) abs s.

exceeds the calibration range by an order of magnitude, owing to an apparent outlier among the series of blank values (Table 2). Therefore the outlier does not represent the ultimately lowest concentration because the noise level was based on only 13 repetitions, which gives a risk of producing an extreme outlier. The LOD of literature values ranges between 3.2×10^{-4} mg dm^{-3} and 5.6 mg dm^{-3}, as dependent on the type of matrix containing the measurment (Table 1), whereas Gill [27] presented a universal LOD of 0.02 mg dm^{-3} for determination of iron by GFAAS. Notably, a limited number of authors report a complete list of figures of merits (Table 1), which may indicate that these numbers are comparable to common values according to the GFAAS technology [27].

3.3.2. Validation of short-term precision

In order to estimate the best possible performance of an analytical apparatus, it is important to perform an operational calibration of the method where standards are analyzed over a short period of time. A short period of time is defined by a single cycle of startup – run – shutdown of the apparatus, as opposed to a long period of time that corresponds to several such independent cycles. This initial exercise gives an estimate of the dilution of samples and expected sensitivity of the analysis. The result of Zeeman GFAAS experiments with analysis of iron by double determinations of standards and a regression line of good quality was obtained as shown in Fig. 1 together with appropriate confidence ranges. It should be noted that the SD of the slope and intercept are relatively small, which indicates that the SD of determination of concentrations of unknowns may be correspondingly small (Fig. 1). Since the SD of the intercept is low and the value of the intercept is large, it may be concluded by t-testing that the intercept is positive and significantly different from zero. This feature attracts immediate attention because there

Table 2. Ten repetitions of blanks for determination of the LOD of iron determination by GFAAS with Zeeman-background correction.

Repetition nr.	Area (Abs·s)
1	0.0125
2	0.0131
3	0.0102
4	0.0124
5	0.0114
6	0.0125
7	0.0129
8	2.8224
9	0.04187
10	0.0347
mean	0.3361
STDEV	0.8828

might be a chemical species present, which interferes with the determination of iron. Therefore this hypothesis was the origin investigated in further detail with respect to magnitude of the non-zero value of intercept.

The figures of merits associated with the initial measurements (Fig. 1) are compatible with those obtained in literature investigations, as shown in Table 1. Although it is generally not recommended [28] as an indicator of quality the correlation coefficient was used as a parameter for estimating the degree of linearity of the regression line, while uncertainties were estimated only in association with determination of the measurand in unknowns (Table 1). An overall uncertainty of samples was reported within the range 1.4–10%, which in many cases is very good precision considering the combined uncertainty of the uncertainty budget encompassing contributions from sample digestions, purities, dilutions, temperature etc. Various magnitudes of intercepts were reported but only in one case had the intercept a value lower than zero following a step of pre-concentration, (Table 1). Thus, it could be anticipated that the general trend was directed towards a positive non-zero value, which corresponds to the findings of Fig. 1.

In the present series of experiments double determinations were performed on samples using the regression line of Fig. 1, which yielded an iron concentration of 29.3±0.3 mg dm^{-1} that was significantly different from the certified reference value of 34.3±1.6 mg dm^{-1}. The uncertainty is represented by the SD of two repetitions which gives a confidence range (95%) of 1.3 mg kg^{-1}. In this case it is not feasible to adjust the uncertainty by the protocol of QUAM [10] using a coverage factor of only two because it is more than an order of magnitude lower than the corresponding *t*-value of 12.7 (N=2, *i.e.*, one degree of freedom). Even after adjustment by the correct t-value, the result of the present analysis was lower than that of the certified value. However, the uncertainty in the measurement was determined on the basis of double determinations but the same uncertainty must be expected to result from an estimate of the uncertainty that was calculated from deviations associated with the regression line. By applying the conventional formula [29] for calculating uncertainty of the sample the uncertainty was determined as 1.1 mg kg^{-1} or more than three times larger than the SD of sample repetitions (Table 1). Therefore a preliminary conclusion was that the measurements were not in statistical control with corresponding uncertainties between repetitions and calibrations. That is, this preliminary result might indicate that the concentration of standards were determined at a different level of uncertainty as compared to the uncertainty ~~on~~ samples.

Although a regression line of good quality supplied a result of low uncertainty (Fig. 1), it should be stressed that the correct value was not reproduced, and this constitutes another cause of concern that prompts of a more thorough investigation of the analytical protocol.

3.4. Pooled calibrations and long-term precision

3.4.1. Lower limit of analysis (LLA)

The lowest value of 0.6 mg dm^{-3} of the LOD obtained in association with the regression line of precision (Fig. 1) does not describe fully the lowest concentration of ordinary analysis because the uncertainty of a concentration becomes large when it is determined close to the LOD. As shown above in the present analysis provides the concept of LOD contradictory results. In addition, the conventional LOD does not take into account the magnitude of residuals of the regression line, which ought to be included in the assessment of a LLA. It is therefore proposed to calculate the LLA that takes into account the SD of intercept (s_{b_0}) and slope (s_{b_1}) under the assumption that the SD of intercept (s_{b_0}) corresponds to the SD of blanks (s_{y_0}) and it is defined as the concentration above which the RSD exceeds 50 % [20]:

$$\mathrm{LLA} = \frac{1}{b_1} \cdot \sqrt{2 \cdot \frac{s_{y_0}^2 + s_{b_0}^2}{1 - 2\left(\frac{s_{b_1}}{b_1}\right)^2}} \cong \frac{1}{b_1} \cdot \frac{2 \cdot s_{b_0}}{\sqrt{1 - 2\left(\frac{s_{b_1}}{b_1}\right)^2}} \qquad (1)$$

RSD > 50%, from pooled calibrations

The approximation of Eq. 1 may be applied when the SD of blanks corresponds to the SD of intercept ($s_{y_0} \cong s_{b_0}$), which is a very good approximation with

pooled calibrations of GFAAS (Fig. 2). Two SDs enter Eq. 1, and it was important to apply a high number of repetitions to construct the regression line for reliably estimating the LLA. An approximate value of reliability (R) as a function of the number of measurements (N) may conveniently be considered by the equation [8]:

$$R = \left[1 - \frac{1}{\sqrt{2 \cdot (N-1)}}\right] \cdot 100\% \qquad (2)$$

Eq. 2 expresses that for example three repetitions, the reliability of the SD becomes 50%, which leaves a high risk for delivering a result that is apparently different from a similar result produced by another laboratory. The present series of experiments produced reliabilities of SDs in the range of 5–83%, which largely depended on the number of repetitions (Table 3). However, if the SD should be determined at a reliability of 95% this would require approximately 200 measurements which usually is too high a number of repetitions for everyday laboratory use, but it would be a number appropriate for validation of methods [30].

3.4.2. Upper limit of analysis (ULA)

Any apparatus of analytical chemistry is supplied with a detector of non-linear response [22,23] but for most practical purposes, the concentration range of linear responses is considered for the sake of simplicity. Despite the simplicity of straight lines, none of the procedures of quality assurance (QA) provides any information about ~~an~~ estimating the full range of concentrations of linear response that applies to operational calibrations. By using the equation for the non-linear response function of detectors it was recently found that a reasonable ULA is found at the concentration that corresponds to a characteristic concentration derived from the expansion to first order of the response function [23]. When the maximum response of the system is denoted as 'A' and the characteristic concentration is denoted as 'B', the response function may be described as:

$$y = A \cdot (1 - \exp[B \cdot (c - c_0)]) \qquad B \leq 0 \qquad (3)$$

where c_0 is the intercept at the abscissa. By expansion of Eq. 3 to first order a straight line of slope '–A·B' and intercept 'A·B·c_0'is obtained but this equation cannot be used as regression line because it represents the tangential line to the response function. The maximum response 'y_{max}' of the tangential line is given by:

$$y_{max,tangential} = A \cdot (1 - \exp[-1]) \qquad (4)$$

which is obtained at the concentration 'c_{max}' that equals the ULA:

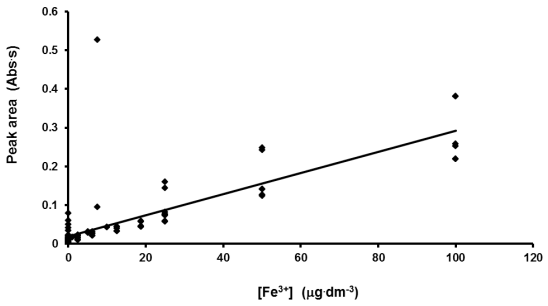

Figure 2. Regression lines for determination of iron by Zeeman GFAAS. All calibration data over several days pooled and depicted in one single diagram without removing outliers.

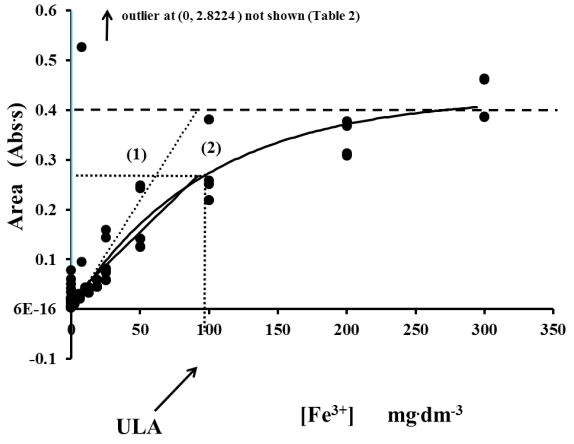

Figure 3. Determination of the upper limit of analysis (ULA) by evaluating responses of standards over the full range of concentrations. The solid curve represents a fitting of Eq. 3 to the experiments and the dotted line at (1) is obtained by expanding Eq. 3 to first order. The straight line at (2) is the regression line that is applicable to calibrations up to concentrations close to the ULA. The horizontal-broken line is the limiting-response value (A) of very high concentrations (Eqs. 3 and 4).

$$ULA = c_{max} = -\frac{1}{B} - c_0 \qquad (5)$$

Thus, two strategies may be chosen for determination of the ULA:

(1) In order to obtain 'A' determine the response value of a standard of very high concentration that is known to provide intensity far above the linear response range. Similar to other procedures it is recommended to switch off the apparatus, switch on again and repeat the cycle as many times as possible. This will provide the more reliable average value for 'A' (Eq. 3).

(2) Perform a series of calibration experiments which exhibit the distinct curvature of Eq. 3 and perform a fitting Eq. 3 to the data thus obtaining ULA directly by the aid of parameters entering Eq. 3.

In some types of apparatus such as mass spectrometers with solid state detectors, it may be impossible to use procedure no. (1) because there is a

Table 3. Analysis by Zeeman GFAAS for iron in four CRM's. Each result was obtained by using a full calibration line for every double determination. The results are grouped according to the concentration of analysis before adjusting for dilution factors. Columns 3-7: All results after measurements at all concentrations. Columns 8-12. Columns 11-14: Results obtained after measurements above the LLA but below the ULA (Fig. 4). Column 13: RSD predicted by Horwitz formula [38].

CRM	Certified value ($mg \cdot kg^{-1}$)	Average value ($mg \cdot kg^{-1}$) All values	STDEV ($mg \cdot kg^{-1}$)	RSD (%)	N	R (%) Eq.2	Average value ($mg \cdot kg^{-1}$) Values between LLA and ULA	STDEV ($mg \cdot kg^{-1}$)	RSD (%)	N	R (%) Eq.2	RSD Horwitz formula (%)
Sea Water	20	36	24	***66***	14	81	-	-	**-**	-	-	***32***
Natural water	34	26	16	***60***	15	81	-	-	**-**	-	-	***33***
Drinking water	650	526	329	***63***	18	83	610	390	***64***	6	68	***29***
Waste water	1480	1211	692	***57***	17	83	945	730	***77***	4	59	***27***

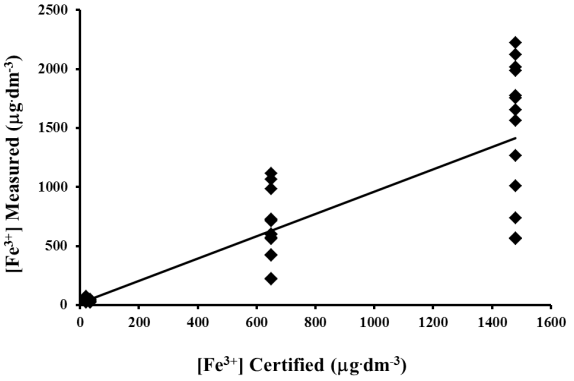

Figure 4. Determination of iron in four different certified reference materials. The determinations were based on results obtained within the range of concentration defined by the LLA and ULA with slope of 0.95 ± 0.09 A s $dm^3 \mu g^{-1}$.

cut-off functionality built into the detection system [23]. In that case option no. (2) may be used to determine 'A' instead of option no. (1). Data with responses lower than or equal to the value given by eq. 4 may be used to construct the regression line. The principle of estimating the ULA is illustrated graphically in Fig. 3. The experiments of Fig. 3 gave a ULA of 97 mg dm^{-3}.

3.4.3. Validation of long-term precision

Calibration may be performed by the manufacturer's software for non-linear-curve fitting following a procedure patented by Unvala [31]. Similar to earlier investigations of the GFAAS technology [20,23] it was found that the SD of responses was almost constant within the full range of calibrations, which seems to be a characteristic property of this particular technology. Since the procedure by Unvala is frequently embedded in the software some of the works copied in Table 1 may have applied the non-linear-working curve for determination of concentrations of unknowns.

It should be noted that the uncertainty of precision, as discussed above, represents the short-term variation of the response of the apparatus while long-term variations of the response was initially not considered. However, long-term variations may conveniently be estimated by repeating measurements in cycles where the apparatus has been shut down completely in between cycles of measurements, as described above. According to GUM [8], the uncertainty of measurements should also be based on previous experiences with the apparatus. Thus, a very important term should be considered in more detail in the uncertainty budget. In Fig. 4 the result of repeating many times the determination of iron concentrations of four different unknowns is shown. The individual measurements depicted in Fig. 4 as a function of consensus values were determined with a lower SD that corresponded to the SD of Fig. 2. Large SDs were associated with determination of iron concentrations in CRMs (Table 3), and the overall RSD was approx. 57-66%, a value almost two orders of magnitude higher than the RSD related to precision (Fig. 1). However, the average value of the CRM was comparable to the certified value with the slope of 0.81 ± 0.09 Abs mg^{-1} dm^3 of the regression line (not shown). The recovery of all results was thus found to be significantly lower than the certified values, which is explained by participation of data that were determined at concentrations far below the LLA. Upon removal of data that were determined at concentrations below the LLA, was the slope of the line changed to 0.95±0.09 Abs mg^{-1} dm^3 (Fig. 4), which shows that the results agreed with certified values. The results of Fig. 4 show that some of the individual data were found close to the value of the CRMs which are 23, 32, 599 and 1268 mg dm^{-3} for the sea water, natural water, drinking water and waste water respectively (Fig. 4 and Table 3). Although the average value was found to correspond to certified values (Table 3), only in rare cases, results were reported with a significant deviation between measured values and certified values [31].

Validation of the method is a key activity in the chain of traceability in analytical chemistry, and method validation is performed most conveniently according to guide of the International Conference on Harmonisation (ICH) [30]. The analytical performance is characterized by figures of merits and by the ability of the method to recover the value of a CRM. Thus, a more detailed method validation was initiated. First, it is proposed that outliers should not be removed from the data set under any circumstance, which is contradictory to the recommendations of International Standardization Organization (ISO) 5725 [11]. Next, a high number of repetitions is required to satisfy the conditions of the central-limit theorem [32]. Despite the proximity of the average values to the certified values, the overall RSD of determinations (Fig. 5) approached 30%. This is an unexpected value in comparison to the prediction of the single regression line of Fig. 1 and to the performances referenced in Table 1. Therefore, in order to understand the unexpectedly high SD in repetitions of unknowns, it was attempted to pool all calibrations of the series of measurements combined [20,21,33]. The result of this presumption is shown in Fig. 2 where data of peak areas are shown as a function of concentrations up to 100 mg dm^{-3}. From this depiction it was found that the RSD depended on concentrations, similar to other findings, and the minimum RSD was found at the ULA of approximately 100 mg dm^{-3} (see above). Apart from the RSDs of very low concentrations, the RSD of calibrations was approximately 30%, which is lower than the RSD of certified references that was determined at concentrations above the LLA (Fig. 4). Accordingly, this result followed the trend also found to exist for other technologies, and it supports the idea of pooling all calibrations into a single series of data that are used for estimating uncertainties but they should not be used for calibration [28]. This procedure thus provides an overview of the true SD that corresponds to the SDs of determination of unknowns. However, it should be stressed that the regression line of pooled calibrations cannot be used to calculate the concentration of unknown. The regression of pooled calibrations is used merely to calculate the standard deviation of calibration. The uncertainty of concentrations based on the regression line may be calculated by the equation [10]:

$$s_c = \left(\frac{1}{b_1}\right) \cdot \sqrt{s_{b_0}^2 + s_y^2 + \left(c \cdot s_{b_1}\right)^2 + 2 \cdot c \cdot \mathrm{cov}(b_0, b_1)} \qquad (6)$$

where b_0 is the intercept, y is the response value of concentration (c), b_1 is the slope and 's' represents the corresponding SDs. However, the covariance term of Eq. 6 may prove slightly difficult to calculate in practice [34], and it was therefore recently suggested that it could be omitted if the remaining terms were divided by the number of terms minus unity. In the case of a straight line was this reduced to division by a factor of two [21]:

$$s_c = \left(\frac{1}{b_1}\right) \cdot \sqrt{\frac{s_{b_0}^2 + s_y^2 + \left(c \cdot s_{b_1}\right)^2}{2}} \qquad (7)$$

Thus, the covariance term as well as the factor of two in the denominator accounts for the fact that it is unlikely that all uncertainties simultaneously adds to the total SD.

In comparison to the reports of Table 1, the uncertainties of the Zeemann GFAAS experiments (Fig. 4) were in some instances more than an order of magnitude higher than expected [35–37]. In order to further assess the validity of the present methodology, the values of the RSDs were calculated by the Horwitz formula [38] (Table 3) and a weak correspondence was found between the Horwitz formula and results obtained between the LLA and the ULA. The results of Table 3 show that the reliability of uncertainty is low when a few repetitions were applied to determine the concentration of unknowns. By introducing the condition that the values should be determined in the calibration range defined by pooled calibrations, that is between LLA and ULA, the number of results was reduced to only a few accepted values (Table 3). In this manner the reliability of those uncertainties decreased to values of 29-69% (Table 3), and the average value of CRMs of low concentrations (20 mg kg^{-1} and 34 mg kg^{-1}, Table 3) were not included because the certified values remained below those of the CRM. The lack of agreement between the RSDs obtained by application of pooled calibrations, and the calculated RSDs on the basis of the Horwitz formula, may result from the influence of interferences on the measured concentrations [20]. However, the number of data is limited, which makes the estimate of uncertainty unreliable. Therefore, the high level of uncertainty of unknowns in table 3 cannot be entirely explained by interferences unless more measurements were included in the investigation. Since the LLA determined by including the extreme outlier of Table 2 was determined as 35.8 mg dm^{-3}, the performance of the present method could not meet the standards of the manufacturer's specifications. Most of the results that were determined by dilution of samples, that is samples of CRMs, were diluted to concentrations below the LLA but this proved to be inapplicable to the present purpose of determination of iron. Therefore, it may be concluded that the parent methodology seems virtually unsuited for determination of iron at the trace-level of concentrations. It should be noted however, that the RSD of samples that were determined without the presence of the extreme outlier was approximately 30% which is in complete agreement

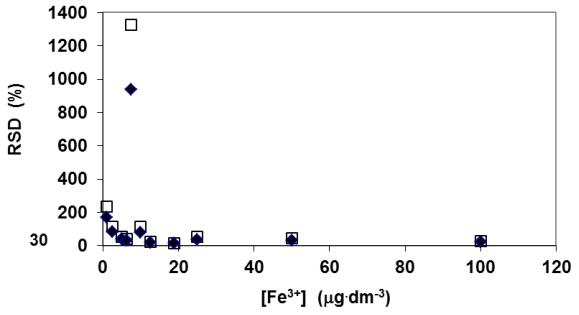

Figure 5. The RSD of calibrations depicted as a function of concentration within the linear range of responses, as calculated by (♦) Peak areas (□) Peak heights. The RSD remained at approx. 30% at concentrations above the LLA and the best choice for analysis was 100 mg dm^{-3} (see text).

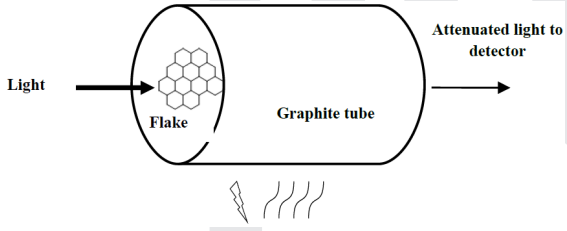

Figure 6. Illustration of the physical mechanism that affects the measured intensity in a manner which is seen as an increase of the magnitude of the standard deviation. Flakes of graphite are liberated during the atomization-step of the graphite furnace. Those graphite flakes, that are detached from the surface of the tube, obstructs by scattering the light at its path towards the detector, gives an apparent absorbance increase in an unpredictable manner, which depends on the size and number of graphite flakes that are released into the light beam.

with calibration-uncertainty and the uncertainty that was predicted by the Horwitz formula. In addition, it was noted that the LLA decreased to approximately 5 mg dm^{-3}, which is a more acceptable value when the outlier was omitted in the calculations. Although it seems to improve the results by excluding outliers, this approach cannot be recommended because the true performance of the apparatus is then hidden, which may cause problems with respect to overall precision. As in most comparable cases, the correct approach is to perform more measurements above the concentration that is defined by the LLA.

3.5. Origin of homoscedasticity

It may be proposed that the origin of this uniform SD of responses across the full range of concentrations is related to solid pieces of graphite that are being liberated from the surface of the graphite tube by firing at high temperatures during measurement. This proposition is illustrated in Fig. 6. Tentatively, the small pieces of graphite liberated from the graphite tube enter the light path and obstruct the light from reaching the detector at all wavelengths, which is thus observed as a high absorbance but in reality it is a matter of scattering. It is well known that the surface of graphite furnaces show changes upon firing [39,40] but the possibility of releasing solid material by influences of temperature and flow of argon gas should also be considered. This suggestion would also explain the high absorbance occasionally observed for blanks (Table 2). It was also noted that the intercept value of all regression lines showed a positive value, while negative values remained absent, and this observation also supports the idea of an additional and uncontrollable contribution by scattered light from solid graphite to the absorbance (Fig. 6). It is possible to extend graphite-tube lifetime by a factor of five by applying tertiary amines reagent to the analysis, which might also diminish the amount of graphite flakes that are released from the tube during analysis [15]. The SD (Eqs. 6 and 7) has the property of approaching a value unique to the parent apparatus and a characteristic value that cannot be altered by anyone using the same apparatus and the same type of sample.

3.6. Determination of s_y

It should be noted that the parameter 's_y' of Eqs. 6 and 7 does not represent the SD on the response of the sample but it represents the SD of a standard with the same concentrations as that of the unknown. It should also be noted that s_y represents the standard deviation of the unknown, as if it were a standard, which thus neglects the influence of potential interferences. It is generally recognized that the value of s_y is approximately proportional to concentration [10] but in the present case of analysis by GFAAS s_y remained constant owing to the high uncertainty generated by the graphite-flake problem (see previous section). It was thus found that s_y = 0.04 ± 0.08 absorbance units. This property is unique to GFAAS with and without Zeeman background correction and it has not been observed earlier, despite its profound influence on SD of the quantitative result [21]. The parameter 's_y' should not be confused with the standard deviation of residuals [29] that enters the IUPAC formula for calculation of standard deviation of the unknown.

3.7. Interferences

Interferences may influence the value of sample concentrations in two different manners (N large). When the determination is in statistical control, corresponds the concentration of unknown exactly to the certified value and s_y(repetition, sample) = s_y(calibration, sample). A potential interference may be identified upon the basis of the deviation from this correspondence:

1. The concentration of unknown corresponds to the certified value and s_y(repetition, sample) > s_y(calibration, sample)

2. The concentration of unknown differs significantly from the certified value

The origin and correction for interferences depends on the type of technology, the sample preparation and on the sampling procedure [41]. Interference due to chloride ions was reported by Volland *et al*. [42], which may explain some of the deviation from the expected value in sea water (Table 3). In order to estimate the magnitude of a potential bias, it is recommended to perform many repetitions and test for differences according to certified values, which may add considerable time to the work with validation. However, once validation is established, the overview of interferences should provide reliable analysis using a relatively low number of samples for future analysis of unknowns. In complicated matrices, such as blood matrices, standards may be prepared with matrix matching, which allegedly minimizes the influence of interferences [43]. However, in the present context, the effect of matrix matching in terms of LLAs (Eq. 1) and SDs (Eq. 7) of pooled calibrations should be investigated. It is well known that determination of concentrations by the aid of standard addition cannot be applied to correct for the influence of interferences, especially when systematic-constant errors are present [44]. Determination of concentrations by standard addition may lead to slightly different results, as shown by Wasilewska [41] but it is a cumbersome task to prove that the observed differences are related to interferences and not to differences in method uncertainties. Since application of standard addition requires detailed knowledge of the linear range of responses, the regression line must be constructed within the full-calibration range. Otherwise, there will be a risk of calibrating within the non-linear range of the response curve, *i.e*., at concentration higher than c_{max} (Eq. 5). Thus, it is rarely favourable to determine concentrations exclusively with standard addition.

4. Conclusion

An investigation of the analytical performance for GFAAS with Zeeman-background correction has provided new knowledge to the field of quality assurance. By using the concept of pooled calibrations, it was found that SDs for response values were equal in the full range of concentrations. This means that homoscedasticity (s_y independent of concentration) was fulfilled for Zeeman GFAAS and it was explained by a release of graphite flakes from the surface of the graphite furnace causing addition of scattering to the absorbance signal during the heating steps of analysis. The present investigation shows that this problem was overlooked in the uncertainty budget of previous investigations.

Generally, outliers should not be removed from any data set. However, in order to obtain unbiased results, the number of measurements to be performed must be very high at low concentrations and correspondingly high SDs may be expected. If a limited number of measurements were performed, it was then found that the SD of determinations increased to high values within the linear range. By evaluating all measurements by pooled calibrations without removing outliers, it was proposed to calculate the LLA and ULA, which is a suitable range for determination of concentrations of unknowns without the need to perform an overwhelming number of repetitions.

By carefully investigating the performance of AAS with Zeeman-background correction, it was found that long-term fluctuations of the detector ought to be included in the evaluation of the precision of measurement. Previously published results indicated that relative uncertainties of Zeeman GFAAS were considerably lower than 10% while the present investigation provided evidence for RSDs that were in excess of 30% in analysis of iron in four different CRMs. This result was provided by that pooling all data of regression lines obtained over a period of several days. If the SDs obtained in this manner were used to estimate the expanded uncertainty, a value in excess of 120% was found, which suggests that many repetitions were required to perform reliable analysis of trace elements with this technology.

Although procedures have been developed for elimination of outliers [2], the present investigation shows that such action will inevitably lead to incomprehensible results and large deviation between predicted SDs and observed SDs. Therefore outliers should not be removed from data sets whatsoever but the influence of outliers on the final result may only be diminished after additional repetitions.

Since the SD of blanks may be extremely high in Zeeman GFAAS it was found that the concept of LOD is a parameter that is inconvenient for characterizing the analytical performance for the technology. Instead it was suggested to use the concept of LLA (Eq. 1) that characterizes the analytical performance at the lowest possible concentration where the RSD is equal to 50%.

It is proposed that all details of the actual validation procedure should be published in future publications that describe new methods of analytical chemistry. This effort ensures that the uncertainty of calibration

corresponds to the uncertainty of repetition, which thus certifies statistical control of the analytical protocol.

The rather large uncertainty observed in the present investigation suggests that the adjustment of uncertainties for uncertainty- budget according to square distributions or triangular distributions of QUAM is virtually unnecessary.

The principle of pooled calibrations, LLA and ULA are very useful concepts introduced for evaluating the analytical performance with respect to long-term precision, and it is proposed that they are applied to all future method-validations of analytical chemistry. These concepts also proved successful for validating of GFAAS with Zeeman-background correction.

A certain degree of correspondence was found between the uncertainty from repetition, uncertainty of calibration and the level of uncertainty predicted by the Horwitz curve. This result indicates that a method validation prepared according to the present protocol provides uncertainties that are comparable to universal uncertainties found by extracting the combined uncertainty from interlaboratory comparisons.

The present methodology provides a means to greatly simplify the work of quality assurance in the laboratory. Although a generally higher level of uncertainty must be accepted, as compared with uncertainties found in contemporary publications, it was suggested that the results were produced with a high degree of reliability.

Acknowledgements

The financial support of the Carlsberg Foundation to establishing the GFAAS with Zeeman-background correction is gratefully acknowledge and many thanks are due to Eva Thale (in memoriam) and to André Markussen for their dedicated technical assistance.

References

1. P. De Bievre PDPT (2000) “Demonstration” vs. “designation” of measurement competence: the need to link accreditation to metrology. Fresenius J Anal Chem 368:567–573.

2. Williams A (2008) Principles of the EURACHEM/CITAC guide “Use of uncertainty information in compliance assessment”. Accreditation and Quality Assurance 13:633–638. doi: 10.1007/s00769-008-0425-3

3. Bednarova M, Aregbe Y, Harper C, Taylor PDP (2006) Evaluation of laboratory performance in IMEP water interlaboratory comparisons. Accreditation and Quality Assurance 10:617–626. doi: 10.1007/s00769-005-0031-6

4. Kreeke J Van De, Calle BD La, Held A, et al. (2010) IMEP-23: The eight EU-WFD priority PAHs in water in the presence of humic acid. TrAC Trends in Analytical Chemistry 29:928–937. doi: 10.1016/j.trac.2010.04.009

5. Papadakis I, Nørgaard J V., Vendelbo E, et al. (2001) International Measurement Evaluation Programme: IMEP-9, trace elements in water. The Analyst 126:228–233. doi: 10.1039/b008619n

6. Voulgaropoulos a., Todorovic M, Cruceru L, et al. (2007) Determination of trace elements in danube water: results of the regional-interlaboratory study. Desalination 213:110–115. doi: 10.1016/j.desal.2006.04.084

7. Resano M, Mozas E, Crespo C, et al. (2010) Solid sampling high-resolution continuum source graphite furnace atomic absorption spectrometry to monitor the biodistribution of gold nanoparticles in mice tissue after intravenous administration. Journal of Analytical Atomic Spectrometry 25:1864. doi: 10.1039/c0ja00086h

8. BIPM I, IFCC I, ISO I (2008) IUPAP and OIML, Evaluation of measurement data-Guide to the expression of uncertainty in measurement. Organization for Standardization (ISO), Online: http://http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

9. Guide C, Edition F (2007) Use of uncertainty information in compliance assessment First Edition 2007.

10. Ellison S, Rosslein M, Williams A (2012) Quantifying uncertainty in analytical measurement. http://www.citac.cc/QUAM2012_P1.pdf

11. (ISO) ISO ISO 5725. http://www.iso.org/iso/home.html. Accessed 30 Apr 2013

12. De Bièvre P (2008) Essential for metrology in chemistry, but not yet achieved: truly internationally understood concepts and associated terms. Metrologia 45:335–341. doi: 10.1088/0026-1394/45/3/011

13. Calle M de la, Kreeke J van de (2008) Report of the third interlaboratory comparison organised by the Community Reference Laboratory for Heavy Metals in Feed and Food. https://www.irmm.jrc.be/EURLs/EURL_heavy_metals/interlaboratory_comparisons/imep-103/Documents/eur23236en.pdf. doi: 10.2787/1332

14. Lin X, Wang L (2010) A semiparametric probit model for case 2 interval-censored failure time data. Statistics in medicine 29:972–81. doi: 10.1002/sim.3832

15. Aleixo PC, Nóbrega J a. (2003) Direct determination of iron and selenium in bovine milk by graphite furnace atomic absorption spectrometry. Food Chemistry 83:457–462. doi: 10.1016/S0308-8146(03)00224-3

16. Bermejo P, Peña E, Domínguez R, et al. (2000)

Speciation of iron in breast milk and infant formulas whey by size exclusion chromatography-high performance liquid chromatography and electrothermal atomic absorption spectrometry. Talanta 50:1211–22.
17. Zödl B, Zeiner M, Paukovits P, et al. (2005) Iron uptake and toxicity in Caco-2 cells. Microchemical Journal 79:393–397. doi: 10.1016/j.microc.2004.10.019
18. Dong X, NakaGuchi Y, Hiraki K (1998) Determination of Chromium, Copper, Iron, Manganese and Lead in Human Hair by Graphite Furnace Atomic Absorption Spectrometry. Analytical sciences 14:785–789.
19. Jakeman a, Thompson T, McHattie J, Lehotay DC (2001) Sensitive method for nontransferrin-bound iron quantification by graphite furnace atomic absorption spectrometry. Clinical biochemistry 34:43–7.
20. Andersen JET (2009) Exercise in Quality Assurance: A Laboratory Exercise. Journal of Chemical Education 86:733. doi: 10.1021/ed086p733
21. Andersen JET (2008) A step-wise approach to the determination of the lower limit of analysis of the calibration line. Journal of Analytical Chemistry 63:308–319. doi: 10.1134/S1061934808040023
22. Andersen JET (2007) Investigation of the performance of the instrument by nonlinear calibration. Microchimica Acta 160:89–96. doi: 10.1007/s00604-007-0804-1
23. Andersen JET (2007) Non-linear calibration leads to improved correspondence between uncertainties. Chemia analityczna 715:715–735.
24. Rorabacher D (1991) Statistical treatment for rejection of deviant values: critical values of Dixon's" Q" parameter and related subrange ratios at the 95% confidence level. Analytical Chemistry 139–146.
25. Huber W (2003) Basic calculations about the limit of detection and its optimal determination. Accreditation and quality assurance 213–217. doi: 10.1007/s00769-003-0626-8
26. Harris DC (2010) Quantitative Chemical Analysis, 8. ed. W.H. Freeman and Company, New York
27. Gill RJ (1993) American Laboratory. American Laboratory 25:F24.
28. Van Loco J, Elskens M, Croux C, Beernaert H (2002) Linearity of calibration curves: use and misuse of the correlation coefficient. Accreditation and Quality Assurance 7:281–285. doi: 10.1007/s00769-002-0487-6
29. Schiller F, Jena D-, Curriet LA (1998) INTERNATIONAL UNION OF PURE COMMISSION ON GENERAL ASPECTS OF ANALYTICAL CHEMISTRY * GUIDELINES FOR CALIBRATION IN ANALYTICAL CHEMISTRY Guideline for calibration in analytical chemistry- Part 1 . Fundamentals and single component. 70:993–1014.
30. Committee IS (1996) ICH Harmonised tripartite guideline, validation of analytical procedures: methodology. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf 1994:
31. Unvala HA (1980) U.S. Patent no. 4.238.830.
32. Geiringer H (1940) A generalization of the law of large numbers. Ann Math Stat 11:393–401.
33. Andersen JET, Mikolajczak M, Wojtachnio-Zawada KO, Nicolajsen HV (2012) The principle of pooled calibrations and outlier retainment elucidates optimum performance of ion chromatography. Journal of chromatography B, Analytical technologies in the biomedical and life sciences 908:122–7. doi: 10.1016/j.jchromb.2012.09.011
34. Raffalt AC (2006) Bestemmelse med ICP-MS af strontium og andre Gruppe II metaller i knogler , knoglemarv og tænder og kvalitetssikring af resultaterne.
35. Patel KS, Shukla a, Goswami a, et al. (2001) A new spectrophotometric method for the determination of total and ferric iron in rain water at the ppb level. Fresenius' journal of analytical chemistry 369:530–4.
36. Soares M, Bastos M, Ferreira M (1993) Determination of copper, iron, aluminium, lead and cadmium in cork stoppers by electrothermal atomic absorption spectrometry. … Analytical Atomic Spectrometry 8:655–657.
37. Von Bonsdorff L, Lindeberg E, Sahlstedt L, et al. (2002) Bleomycin-detectable iron assay for non-transferrin-bound iron in hematologic malignancies. Clinical chemistry 48:307–14.
38. Horwitz W (1982) Evaluation of Analytical Methods Used for Regulation of Foods and Drugs. Analytical Chemistry 54:67A–76A. doi: 10.1021/ac00238a765
39. Ashino T, Shimabukuro H, Wagatsuma K (2010) Degradation Process of Graphite Furnace Estimated from the Atomic Gas Temperature of Iron in Graphite Furnace Atomic Absorption Spectrometry. Materials Transactions 51:2215–2219. doi: 10.2320/matertrans.M2010180
40. Hilligsøe B, Andersen J.E.T., Hansen E.H. (1997) Investigations into the role of modifiers for entrapment of hydrides in flow injection hydride generation electrothermal atomic absorption spectrometry as exemplified by. Journal of Analytical Atomic Spectrometry 12:585–588.
41. Wasilewska M (1997) Influence of iron on Bisphenol A quality. Spectrochimica Acta Part B: Atomic Spectroscopy 52:1321–1326. doi: 10.1016/S0584-8547(97)00043-8
42. Volland G., Tschöpel P. TG (1977) Electrolytische Abscheidung im Hydrodynamischen System von ng-mengen Eisen, Kobalt, Zink und Wismut im Graphitrohr. Analytica Chimica Acta 90:15–23.
43. Bannon DI, Murashchik C, Zapf CR, et al. (1994)

Graphite furnace atomic absorption spectroscopic measurement of blood lead in matrix-matched standards. Clinical chemistry 40:1730–4.
44. Baraj B, Bianchini A, Niencheski LFH, et al. (2001) The Performance of Zeiss GFAAS-5 Instrument on the Determination of Trace Metals in Whole Blood Samples of Southern Elephant. Fresenius Environmental Bulletin 10:859–862.